The Performance of the NAS HSPs in 3Q93

Robert J. Bergeron

Report RND-009 August 1994

NAS Systems Development Branch
NAS Systems Division
NASA Ames Research Center
Mail Stop 258-6
Moffett Field, CA 94035-1000
bergeron@nas.nasa.gov

Abstract

During 3Q93, the NAS C-90 delivered an average throughput of 3.442 GFLOPS while the NAS Y-MP averaged 0.705 GFLOPS. A 3% greater system availability permitted the third quarter C-90 throughput to exceed that of the second quarter by 3%. Although 3Q93 vector hardware length and vector operation fraction were slightly less than that of 2Q93, reduced overhead in 3Q93 allowed an increased C-90 CPU performance.

1.0 Introduction

The introduction of the C-90 in March 1993 motivated the daily monitoring of the hardware performance of the NAS High Speed Processors (HSPs). This paper, covering the 3rd quarter of 1993, is the second report in the series[1].

The C-90 Hardware Performance Monitor (HPM) continuously delivers a full 32-counter record for the workload and the C-90 discussion will include a report of all 32 counters. The Y-MP HPM can report only a single group (8-counter) during any one period[2]. Since NAS has chosen to monitor the Group 0 performance of all programs in the Y-MP workload, the Y-MP discussion will include only the Group 0 counters. This group provides a general quantitative overview of program performance.

NAS records the daily average values of all HPM counters. In 3Q93, the C-90 memory upgrade from 256 WM to 1 GW in late September permitted 88 days of measured user service. This report provides tables of counter values representing the average, maximum, and minimum values from the 88 daily reports in this quarter. The Y-MP counters represent 90 days of user service.

The tables provide performance rate data per CPU for the actual time the CPU spent executing the user programs. System throughput, however, is based on wall clock time and total number of CPUs. A complete explanation of all counter data occurs in [3].

To provide a feel for the daily variation in each of the counters, the report also provides the standard deviation (STD) and coefficient of variation (COV). The coefficient of variation is the ratio of the standard deviation of a quantity divided by its average value.

2.0 C-90 Counter Data

The C-90 CPUs have a clock period (CP) of 4.167 nanoseconds and a peak speed of 960 MFLOPS. This architecture has 128-element vector registers and double-width 64-element functional units. At the end of the second month of 3Q93, the C-90 experienced a substantial increase in memory from 256 MW to 1 GW and the end of this section presents some comments on the initial impact of this upgrade.

For clarity, the report divides the 32 C-90 counters into 4 functional groups: global counters, instruction holds, instruction issues, and vector operations.

Global Counters:

Table 1 provides counter data giving a total counts for instructions, operations and references. The unit "M/sec" denotes "Million per sec" and the unit "avg/ref" denotes "average (conflict) per reference". The term "reference" denotes a single Cray word (8-byte) data transfer.

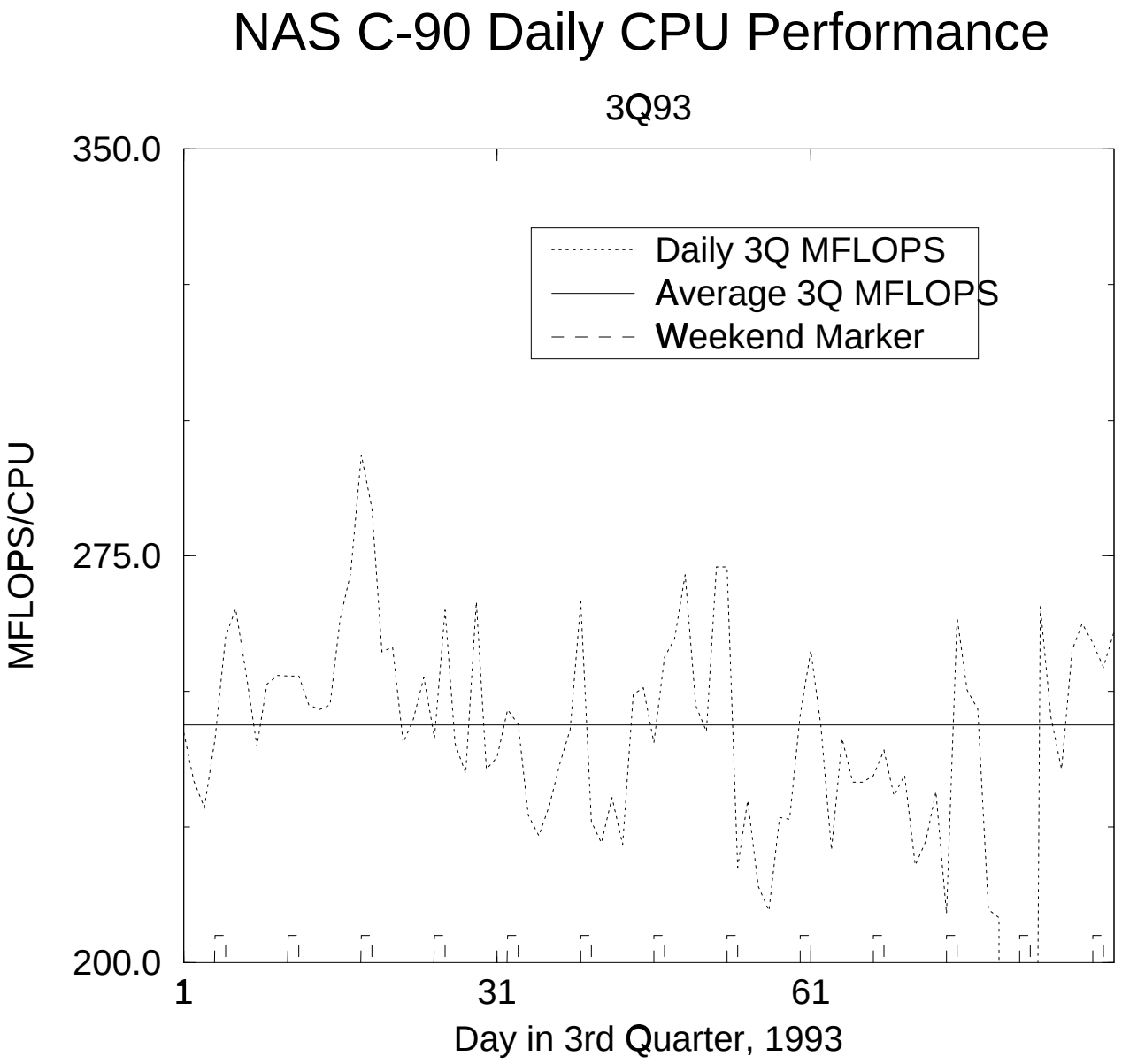
Table 1: NAS C-90 3Q93 Daily Average HPM Measurements- Global Counters

Measurement	Unit	Avg	STD	COV	Min	Max
CPU time	sec	62166.	18630.	0.299	15951.	80631.
Instruction Issue	M/sec	59.870	3.268	0.055	50.434	69.210
Average clock periods/inst	--------	4.020	0.222	0.055	3.467	4.758
CP holding issue	Percent	66.158	1.837	0.028	60.468	71.550
Instruction buffer fetches	M/sec	0.270	0.027	0.099	0.205	0.331
Floating Pt. Ops per CPU	M/sec	243.913	17.254	0.071	208.273	293.675
Vector Floating Pt. Ops	M/sec	239.199	17.741	0.074	200.463	290.326
CPU memory references	M/sec	231.857	16.295	0.070	186.723	275.219
CPU memory conflicts	avg/ref	0.275	0.031	0.112	0.203	0.374
VEC memory references	M/sec	226.359	16.774	0.074	179.511	271.457
B/T memory references	M/sec	1.333	0.228	0.171	0.842	2.150
I/O memory references	M/sec	3.170	1.317	0.416	0.399	6.266
I/O memory conflicts	avg/ref	0.282	0.018	0.063	0.234	0.322

The large variation in the CPU time measurement reflects the requirement that the HPM data represent a continuous interval. Occasionally, persistent hardware and/or software problems may require several shutdowns during the 24-hour measurement period. The CPU time reported in the table for such days is the longest continuous period without a shutdown.

During 3Q93, each of the C-90's 16 CPUs performed at an average rate of 244 MFLOPS, with the vector units contributing 98% of the floating point operations. The CPU MFLOP rate exceeds the CPU memory reference rate by a factor of 1.05, indicating that the average floating point operation must be reusing data in the registers to avoid memory accesses.

During the weekend periods, CPU-intensive batch jobs dominate execution time and produce a somewhat higher performance rate than those of the weekdays. The weekend average is 260 MFLOPS, whereas the weekday average is 242 MFLOPS. The following figure shows the average daily CPU performance rate during 3Q93. Since performance peaks occur in both weekends and weekdays, the quality of the batch submittals must be similar for both periods. Installation of the 1 GW memory near the end of the 3Q93 accounted for the drop in performance shown in the figure.



The I/O memory reference rate of 3.17 Mwords/sec per CPU represents an average sustained I/O rate of 0.406 Gbytes for the machine. This rate includes I/O which is both internal and external to the machine. Since the SSD can sustain 13.6 Gbytes/sec and the 68 DD60s can sustain 1.4 Gbytes/sec, the C-90 I/O does not, on the average, challenge the data transfer capacity of its disks.

The I/O rate measurements display a large COV relative to the performance rate measurements.This variance reflects the differing input/output requirements of NAS users. The Cray timesharing architecture decouples the I/O rate from the MFLOP rate because the data transfer occurs when the user program has given up control of the CPU to another program. The second program can maintain the CPU MFLOP rate while the I/O from the first program proceeds. If the transfer is efficient and the two programs have similar performance characteristics, measurements should show the MFLOP rate relatively constant while the I/O rate fluctuates according to user needs. The C-90 measurements substantiate this claim.

Instruction Holds:

The instruction processor fetches instructions from the instruction buffer. If any of the resources required to execute the instruction are reserved, the instruction issue logic prevents the instruction from issuing. The HPM records all CPs for which the instruction holds issue and Table 2 presents these as the percent of total CPU time. Since there may be more than one resource reservation preventing an instruction issue, the sum of the percentages in this group can exceed 100%.

Table 2: NAS C-90 Daily Average 3Q93 HPM Measurements-Instruction Holds

Measurement	Unit	Avg	STD	COV	Min	Max
Waiting on A-registers	% CPU	4.891	0.341	0.070	3.816	5.782
Waiting on S-registers	% CPU	9.312	1.327	0.143	5.712	13.310
Waiting on V-registers	% CPU	21.727	1.740	0.080	17.534	28.953
Waiting on B/T-registers	% CPU	1.226	0.183	0.149	0.858	1.914
Waiting on F'nctnal Units	% CPU	23.053	1.326	0.058	19.574	26.727
Waiting on Shared Regs	% CPU	0.456	0.288	0.632	0.008	1.535
Waiting on Memory Ports	% CPU	16.584	1.720	0.104	11.671	22.390
Waiting on Miscellaneous	% CPU	2.386	0.122	0.051	2.113	2.663

For the NAS C-90 workload, the major resources causing instruction issue delays are busy vector registers and busy vector functional units. The approximately equal delays in vector registers and vector functional units indicates efficient register use and overlapping of vector functional units.

Memory references can lead to two kinds of delay in the Y-MP/C-90 architecture. A memory instruction hold occurs, for example, when a register is reserved by another instruction or a memory port is busy. Table 2 shows that the fraction of CPU time the processor held issue is about 17%.

The second type of memory delay is termed a memory conflict (or memory contention), and this delay occurs when a needed bank is busy. A CPU memory port accesses a section which accesses a memory bank. A user program executing on a single CPU can encounter conflicts when it continuously references the same bank. A workload can encounter conflicts when several CPUs simultaneously reference the same bank.

The total delay due to memory references includes both the delays due to memory contention as well as the delays due to memory instruction holds. Data from Table 1 indicates that each memory reference on the average experiences a memory contention delay of 0.275 CP. Data from Table 2 indicates that memory references prevent the CPU from issuing an instruction about 17% of the time.

The memory instruction issue delay expressed on a per reference basis is the ratio of number of clock periods delayed divided by the number of memory references:

where

$$InstructionIssueDelay = \frac{0.16584 \times 62166 \times 2.3998e08}{231.857e06 \times 62166} = 0.172$$

0.16584 = fraction of CPU time spent waiting on memory ports (Table 2),

62166 = CPU seconds measured in 3Q93 (Table 1),

2.3998e08 = Number of C-90 CPs in one second, and

231.857e06 = Average number of C-90 CPU memory references per second (Table1).

The total memory delay is the sum of the memory contention delay and the memory instruction issue delay or (0.275 + 0.172) CP per reference. For the measured vector lengths of 63 (Table 4), the Y-MP can load data from memory at a rate of 1.30 CP/word; the Y-MP can store data to memory at a rate of 1.05 CP/word. C-90 rates are expected to be similar. Total memory delay is about 0.45 CP/reference. Since this delay is a fraction of the 1.05 CP minimum required for a C-90 workload vector memory reference, memory is not a bottleneck for the C-90 workload at the current vectorization level.

Instruction Issues:

Instructions produce the operations which constitute the actual workload tasks.

Table 3: NAS C-90 Daily Average 3Q93 HPM Measurements-Instruction Issues

Measurement	Unit	Avg	STD	COV	Min	Max
(000-004)Special	M/sec	1.225	0.148	0.121	0.830	1.613
(005-017)Branch	M/sec	2.541	0.218	0.086	1.969	3.122
(02x,030-033)A Register	M/sec	25.613	1.774	0.069	20.383	30.769
(034-037)B/T Memory	M/sec	0.150	0.031	0.205	0.082	0.280
(040-043,071-077)S Register	M/sec	8.283	1.114	0.134	5.393	12.357
(044-061)Scalar Integer	M/sec	4.433	0.600	0.135	2.998	6.619
(062-070)Scalar Floating Pt.	M/sec	4.714	0.965	0.205	2.500	8.623
(10x-13x)Scalar Memory	M/sec	4.166	0.586	0.141	2.788	6.160
(140-177)All Vector	M/sec	8.745	0.567	0.065	7.104	10.046

A-register instructions comprise about 43% of the scalar instructions issued. These instructions compute memory addresses and indexes for memory, loop control, and I/O. The C-90 percentage of A-register instructions exceeds the 32% reported for the NAS Y-MP workload[3] and this increase appears due to the additional address overhead of the multiple pipes. All CPUs of the C-90/Y-MP architecture have two pipes, one consisting of an add functional unit and the other consisting of a multiply functional unit. In both architectures, the number of segments in the functional unit is 64. However, the C-90 functional unit design provides two 64-element columns and these double-width functional units require some additional A-register operations.

Scalar instructions constitute about 36% of NAS workload instructions. About 38% of these scalar instructions employ the S-registers for logical functions and special data transfers. Integer, floating point, and memory instructions provide approximately equal contributions to the remainder of the scalar instructions.

Vector instructions are only 15% of the total instructions, but vector operations represent about 91% of the workload operations (Derived Data). A single vector instruction can produce up to 128 vector operations.

Vector Operations:

All of the vector operations shown in Table 4 are produced by vector instructions and in Table 3, the rate of instruction issue for the vector instructions was 8.745 million per second. The vector operation rate for 3Q93 was 520 million per second.

Table 4: NAS C-90 Daily Average 3Q93 HPM Measurements- Vector Operations

Measurement	Unit	Avg	STD	COV	Min	Max
Vector Logical	M/sec	29.072	3.818	0.131	20.895	38.844
Vector Shift/Pop/LZ	M/sec	8.926	1.266	0.142	6.700	14.200
Vector Integer Add	M/sec	16.904	2.510	0.148	11.736	24.537
Vector Floating Pt. Multiply	M/sec	121.304	9.453	0.078	99.396	149.798
Vector Floating Pt. Add	M/sec	111.328	8.656	0.078	94.557	132.493
Vector Floating Reciprocal	M/sec	6.567	0.896	0.136	4.817	8.759
Vector Memory Read	M/sec	157.489	11.868	0.075	124.673	191.779
Vector Memory Write	M/sec	68.869	5.642	0.082	54.838	89.607
Average Vector Length	--------	59.687	4.554	0.076	48.330	71.700

The workload vector length is about 60 whereas the C-90 hardware vector length is 128. The measured value is less than the program vector length. For example, a program with a logical vector length will 192 require two vector instructions, one for the 128 element block and a second for the 64 element block. The HPM will report a vector length of about 96 for this example.

Vector memory load (read) rates are twice as large as vector memory store (write) rates. A FLOP requires, on the average, one memory reference, but the NAS data indicate that it is more likely to be a load than a store. The C-90 architecture provides each CPU with two memory double-width paths for loading data from memory and one double-width memory path for storage; the architecture reserves the fourth memory path for I/O and instruction buffer transfers[4]. The current workload displays a maximum CPU memory bandwidth of 1.2 references per CP whereas the C-90 provides a maximum memory bandwidth of 6 references per CP per CPU.

Derived Data:

The table lists several quantities obtained through calculations with the counter data.

Table 5: NAS C-90 Daily Average 3Q93 HPM Measurements-Derived Data

Measurement	Unit	Avg	STD	COV	Min	Max
System Availability	Percent	88.30	4.90	0.055	70.00	94.00
System MFLOPS	M/sec	3442.414	290.239	0.084	2643.000	4327.270
Vector Operation Fraction	Percent	91.007	0.969	0.011	87.690	93.410
Scalar Operation Fraction	Percent	8.993	0.969	0.108	6.590	12.310
Vector Operation Rate	M/sec	520.459	33.922	0.065	432.510	617.510
Scalar Operation Rate	M/sec	51.125	3.207	0.063	42.110	60.700
Total Operation Rate	M/sec	571.582	31.654	0.055	493.210	662.950
Instruction Issue Fraction	Percent	24.947	1.358	0.054	21.016	28.840
Hold Issue Fraction	Percent	66.158	1.830	0.028	60.468	71.550
Null Instruction Fraction	Percent	8.895	0.613	0.069	7.190	10.692

Availability is the fraction of time the C-90 operated in user mode. During other times, the C-90 was either idle or executing system calls. The low average availability is generally due to kernel conflicts, i.e., the CPUs performing system activity must wait while a single CPU updates critical kernel tables. The 8.0 version of the UNICOS operating system (available 12/93) will allow multiple CPUs to modify kernel data to reduce system time generated by this problem.

System MFLOPS denotes the system throughput. This rate is the product:

System MFLOPS = (MFLOPS/CPU) *CPUs*Availability.

The table shows the throughput rate to be 3442 MFLOPS or 22.4% of peak.

The table indicates that 91.0% of the operations were performed in vector mode and a total operation rate of 572 MOPS. Since this rate is about 2.38 OPS/CP, the instruction processor is able to overlap operations despite the large number of hold issue CPs.

Of the 572 MOPS, 40% were memory operations. If typical machine operations followed the “Load,Load,Operate, and Store” pattern, 75% of the operations would be memory operations. The reduced amount of memory usage confirms that the compiler is successfully using registers to limit memory operations.

A complete accounting of all CPs accumulated by the C-90 CPU while in user mode includes the time spent issuing instructions, the time spent holding instruction issue, and the time spent preparing for the next instruction. Cray terms the latter quantity NIP (Next Instruction Parcel) time and includes in it the CPs spent jumping across instruction buffers, CPs spent fetching words from memory to load the buffers, and CPs spent processing instructions of more than one word in length. Partition of total CP time into these categories can illustrate the reasons for performance differences between two similar workloads. In 3Q93, the C-90 spent about 25% of the user time issuing instructions, about 66% of the user time holding issue, and about 9% of the user time preparing for the next instruction.

The 1 GW upgrade in the C-90 memory contributed little to the third quarter performance since it occurred very late in the quarter. Comparison of the post-upgrade performance at the time of this writing (1 month after the upgrade) to the pre-upgrade performance indicates a substantial(12%) increase in the vector length and a smaller (3%) increase in CPU MFLOP rate.

3.0 Y-MP Counter Data

The Y-MP CPUs have a clock period (CP) of 6.000 nanoseconds and a peak speed of 333 MFLOPS. Table 6 provides 3Q NAS Y-MP per-CPU counter data for Group 0. Table 7 provides values calculated from the counter data.

Group 0:

Table 6: NAS Y-MP Daily Average 3Q93 HPM Measurements-Group 0

Measurement	Units	Average	STD	COV	Min	Max
CPU time	Sec	72974.	12556.	0.172	26376.	81930.
Instruction Issue	-----	36.417	2.646	0.073	30.950	42.740
Average clock periods/inst	M/sec	4.601	0.334	0.073	3.900	5.380
CP holding issue	Percent	68.627	2.440	0.036	63.540	73.980
Instruction buffer fetches	M/sec	0.295	0.036	0.120	0.200	0.380
Floating Pt. adds	M/sec	45.476	3.438	0.076	37.650	53.280
Floating Pt. multiplies	M/sec	49.251	3.726	0.076	42.310	56.920
Floating Pt. reciprocals	M/sec	3.230	0.345	0.107	2.410	4.220
CPU memory references	M/sec	99.797	7.451	0.075	82.930	114.850
I/O memory references	M/sec	0.490	0.345	0.705	0.100	2.620
Floating Pt. Ops per CPU	M/sec	97.957	7.136	0.073	83.660	112.630

The Group 0 counters provide total counts of instructions, operations and memory references. Reasons for the large COVs for CPU time and I/O memory reference rate are the same as those provided for the C-90 in Section 2.

During 3Q93, each of the Y-MPs 8 CPUs performed at an average rate of 98 MFLOPS. The CPU MFLOP rate is slightly less than the memory reference rate, indicating that each floating point operation requires slightly more than one memory reference. Since the value of the ratio of FLOPS to memory references is close to unity, the compiler must be reusing registers to avoid main memory access.

The I/O memory reference rate of 0.490 Mwords/sec per CPU represents an average sustained I/O rate of 0.031 Gbytes for the machine. This rate includes I/O which into both internal and external to the machine. For rates internal to the machine the relevant targets are again the SSD and the disks in the 2 IOSs (68 DD40s). Since the SSD can sustain 1.6 Gbytes/sec and the 40 DD40s and 8 DD49s can sustain 0.460 Gbytes/sec, the Y-MP I/O does not, on the average, challenge the data transfer capacity of its disks.

Derived Data:

Table 7 lists several quantities obtained through calculations with the counter data.

Table 7: NAS Y-MP Daily Average 3Q93 HPM Measurements-Derived Data

Measurement	Units	Average	STD	COV	Min	Max
System Availability	Percent	89.9	3.8	0.043	72.0	95.0
System MFLOPS	M/sec	705.313	67.172	0.095	524.420	842.720
Vector Operation Fraction	Percent	86.				
Scalar Operation Fraction	Percent	14.				
Vector Operation Rate	M/sec	201.				
Scalar Operation Rate	M/sec	33.				
Total Operation Rate	M/sec	234.68				
Instruction Issue Fraction	Percent	21.825	1.583	0.072	18.570	25.644
Hold Issue Fraction	Percent	68.527	2.440	0.036	63.540	73.980
Null Instruction Fraction	Percent	9.547	0.963	0.101	7.446	11.644

Availability is the fraction of time the Y-MP operated in user mode. During other times, the Y-MP was either idle or executing system calls.

System MFLOPS denotes the system throughput. This rate is the product:

System MFLOPS = MFLOPS/CPU *CPUs*Availability.

The table shows the throughput rate to be 705 MFLOPS or 26.5% of the theoretical peak rate.

The table provides a value of 86% for the fraction of Y-MP operations performed in vector mode based on previous NAS measurements [3].

The total operation rate includes the operations produced by the vector and the scalar instructions. The Y-MP Group 0 output does not provide vector and scalar integer and logical operations. The value in Table 7 is the sum of the Group 0 memory operation rate, floating point operation rate, and instruction issue rate. Experience indicates that the latter can serve as a rough approximation to the logical and integer operations not included in Group 0. The aggregate rate of 235 MOPS per CPU represents about 1.40 OPS/CP, indicating that the instruction processor is able to overlap operations despite the large number of hold issue CPs.

Of the 234 MOPS, 43% were memory operations. If typical machine operations followed the "Load, Load, Operate, and Store" pattern, 75% of the operations would be memory operations. The reduced amount of memory usage confirms that the compiler is successfully using registers to limit memory operations.

As with the C-90, a complete accounting of all CPs accumulated by the CPU while in user mode includes the time spent issuing instructions, the time spent holding instruction issue, and the time spent preparing for the next instruction. The Y-MP spent about 23% of the user time issuing instructions, 67% of the user time holding issue, and 10% of the user time preparing for the next instruction. The Next Instruction Fraction, while still constituting a small amount of user CPU time is about 10% larger than that of the C-90.

4.0 Discussion

The 3Q C-90 average CPU performance was about the same as that of the previous quarter, even though the percent vectorization and vector length were less than those of 2Q93. The 3Q93 user overhead, defined as Next Instruction Packet (NIP) time was somewhat less than 2Q93, i.e., 8.935% vs. 9.235%. The smaller overhead enabled the C-90 CPUs to display an average performance similar to that of 2Q93. The smaller overhead also manifests itself in the issue rate for special and branch instructions and in the vector logical operation rate. The 3Q93 measurements indicate a decrease in all of these rates relative to 2Q93.

The C-90 throughput, defined as

System MFLOPS = (MFLOPS/CPU) *CPUs*Availability

was about 3% larger than that of the previous quarter and a 3% increase in the availability was the reason for the increase. An increased C-90 usage in 3Q93 apparently did not lead to increased system and idle times.

The 3Q93 Y-MP average CPU performance was about the same as that of the previous quarter and, as with the C-90, a slightly higher system availability produced an increased throughput in 3Q93.

Peak performance requires asymptotically long vector lengths to reduce the vector start-up penalty to zero. Although the users are employing similar codes on both machines, the C-90 user programs display vector lengths which are short relative to the longer C-90 hardware length. These programs do not fully exploit the vector hardware and lead to a lower efficiency relative to the Y-MP, 22.1% vs. 26.4%. HPM measurements in the 10-day interval following the introduction of the 1 GW memory indicated a 12% increase in vector lengths. Measurements have not indicated an increase in memory instruction holds due to the larger memory, but future reports will monitor this quantity closely.

The 3Q C-90 CPU I/O rate exceeded the Y-MP CPU I/O rate by a factor of 6.5 whereas the CPU MFLOP rate exceeded the Y-MP MFLOP rate by 2.5. In 2Q93, the I/O ratio was 3.2 whereas the CPU ratio was 2.5. While the 3Q C-90 I/O rate was somewhat higher than expected, the measurements indicate the increased I/O did not detract from CPU performance.

5.0 Conclusion

During 3Q93, the C-90 sustained a 3.4 GFLOP system throughput rate on its NAS workload while the Y-MP maintained a 0.7 GFLOP rate on essentially the same NAS workload.

Although some numerical algorithms in the workload, such as multigrid and sparse matrix solvers, display inherently short vector lengths, the NAS Technical Summaries indicate these algorithms form a minority of workload programs. While the 2Q93 HPM measurements disclosed no obvious resource bottlenecks, the average vector length needs to increase to fully utilize the double-width functional units. While memory utilization for both machines was well over 90%, the current NAS workload does not challenge the computing power of the C-90 because the workload vector lengths are less than half of the hardware maximum.

Although the NAS workloads strongly underutilize the massive I/O capacity of both machines, individual users employing the SSD may be making good use of the Cray I/O capability.

6.0 Acknowledgment

Thanks to David Barkai and Toby Harness for reviewing this paper.

7.0 References

[1] Bergeron, R.J."The Performance of the NAS HSPs in 2Q93",NASA Ames Research Center, NAS Systems Division Report RND-94-009, January, 1994.

[2] Cray Research, Inc. UNICOS Performance Utilities Reference Manual, SR-2040 7.0, 1992.

[3] Bergeron, R.J."Performance Analysis of the NAS Y-MP Workload", NASA Ames Research Center, NAS Systems Division Report RND-90-009, December, 1990.

[4] Cray Research, Inc. CRAY Y-MP System Programmer's Reference Manual, CSM-04--0A0, 1989